Feature-Space Separation and Classification of Heart–Lung Audio from Mixed Single-Channel Recordings

Filipi Kunz Alexander Yoon Austin Holt
Georgia Institute of Technology

December 2025

Abstract

Digital stethoscope recordings frequently contain simultaneous heart and lung sounds, creating challenges for downstream classification and limiting clinical interpretability. In this project, we construct a full machine-learning pipeline that (1) loads, normalizes, and explores heart, lung, and mixed cardiopulmonary recordings; (2) extracts robust acoustic features; (3) performs supervised source separation in *feature space* using multi-output Random Forest regression; and (4) classifies predicted heart and lung features into clinically meaningful labels. Using the HLS-CMDS dataset, which provides mixed recordings with matched heart-only and lung-only references, we validate the pipeline through cross-validation, bootstrap confidence intervals, and detailed regression diagnostics. We additionally provide a rigorous shortcomings analysis covering dataset, methodological, statistical, and generalization limitations.

1 Introduction

Digital auscultation is central to cardiopulmonary assessment. Modern stethoscopes record high-resolution audio, enabling automated diagnostic approaches. However, chest recordings typically contain overlapping acoustic events from both the heart and lungs. Heart sounds (S1, S2, murmurs) occupy low frequencies (20–150 Hz), whereas lung sounds (wheezes, crackles, airflow turbulence) span broad higher frequencies (100–1000+ Hz). When these sources overlap in a single-channel recording, direct interpretation becomes difficult.

The project proposal (included in the submission as :contentReference[oaicite:2]index=2) outlines the motivation for a unified analysis framework that separates heart and lung signals and produces reliable classification outputs. Furthermore, the dataset README (:contentReference[oaicite:3]index=3) describes a unique dataset in which mixed recordings have matched reference heart-only and lung-only waveform files. This allows supervised models to learn mappings that would otherwise require blind source separation.

Traditional separation techniques such as Nonnegative Matrix Factorization (NMF) [3], Independent Component Analysis (ICA) [4], and Wiener filtering [5] tend to fail when applied to cardiopulmonary recordings due to:

- overlapping frequency bands,
- lack of statistical independence between sources,
- single-channel mixture structure.

Instead of waveform separation, we pursue a *feature-space regression* approach:

1. Extract a feature vector from the mixed signal.
2. Train supervised regressors to predict the heart-only and lung-only feature vectors.
3. Classify the predicted features for each organ.

This approach sidesteps the challenges of waveform decomposition, yet still allows evaluation against real reference signals.

The rest of this report follows the required structure: 1) Introduction, 2) Methods, 3) Results, 4) Conclusion, 5) References.

2 Methods

2.1 Dataset

We use the Heart and Lung Sounds Dataset Recorded from a Clinical Manikin Using a Digital Stethoscope (HLS-CMDS). The dataset consists of:

- 50 heart-only recordings,
- 50 lung-only recordings,
- 145 mixed recordings (each with multiple filtered variants),
- reference heart and lung tracks corresponding to each mixture.

Audio is 22,050 Hz PCM WAV, 15 seconds per file. Metadata includes:

- chest location (e.g., apex, base, anterior, posterior),
- simulated gender,
- sound type (normal, wheeze, crackle, abnormal rhythm).

Because mixed recordings come with matched references, we can train supervised regressors for source separation.

2.2 Stage 1: Loading, Normalizing, EDA

Recordings are loaded using a standard WAV reader and normalized to $[-1, 1]$. We validate:

- consistent duration across files,
- lack of clipping,
- presence of metadata,
- distribution of gender, sound type, and location.

We produce (not shown here):

- histograms of metadata distributions,
- correlation heatmaps,
- spectrogram comparisons (STFT).

2.3 Stage 2: Feature Engineering

We compute a comprehensive set of features inspired by speech processing [6], biomedical acoustics [7], and previous cardiopulmonary analyses.

2.3.1 Spectral Features

Using short-time Fourier transform (STFT) windows, we compute:

- spectral centroid,
- spectral bandwidth,
- flatness,
- rolloff at 85% energy,
- harmonicity metrics.

2.3.2 MFCCs

Following [6], we compute 13 MFCCs and their first time derivatives.

2.3.3 Temporal Features

- RMS energy,
- zero-crossing rate,
- amplitude envelope statistics.

2.3.4 Organ-Specific Features

Heart: S1 and S2 energy ratios, beat-to-beat interval variance. **Lung:** wheeze-band energy (300–800 Hz), crackle transient counts.

All features are concatenated into a single vector (dimension ≈ 150).

2.4 Stage 3: Feature-Space Supervised Separation

Let:

$$X_{\text{mix}} \in \mathbb{R}^d, \quad Y_{\text{heart}}, Y_{\text{lung}} \in \mathbb{R}^d.$$

We train multi-output Random Forest regressors:

$$\hat{Y}^{(H)} = f_H(X_{\text{mix}}), \qquad \hat{Y}^{(L)} = f_L(X_{\text{mix}}).$$

2.4.1 Random Forest Regression

A Random Forest regressor [8] consists of T decision trees. Each tree produces prediction $h_t(x)$. The ensemble output is:

$$\hat{y}(x) = \frac{1}{T} \sum_{t=1}^{T} h_t(x).$$

Trees are trained on bootstrap samples with random feature subsets, lowering variance relative to a single tree.

2.4.2 Bias–Variance Decomposition

Prediction error decomposes as:

$$\mathbb{E}[(\hat{f}(x) - f(x))^2] = \underbrace{\left(\mathbb{E}\hat{f}(x) - f(x)\right)^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}\left[(\hat{f}(x) - \mathbb{E}\hat{f}(x))^2\right]}_{\text{Variance}} + \sigma^2.$$

Random Forests primarily reduce variance via averaging.

2.5 Stage 4: Classification

Using predicted features, we classify each organ's sound type.

Models evaluated:

- Random Forest,
- Logistic Regression,
- k-Nearest Neighbors,
- Gaussian Mixture Models.

2.5.1 Cross-Entropy Loss

For prediction $p(y = k|x)$:

$$L_{\text{CE}} = -\sum_{i=1}^{n} \sum_{k=1}^{K} \mathbf{1}(y_i = k) \log p(y_i = k).$$

2.6 Evaluation: Cross-Validation and Bootstrapping

We use 5-fold cross-validation. Additionally, we compute bootstrap confidence intervals for all metrics.

2.6.1 Bootstrap CI

Given bootstrap statistics $\{\hat{\theta}_b\}_{b=1}^{B}$, the percentile interval is:

$$\left[\hat{\theta}_{(\alpha/2)}, \hat{\theta}_{(1-\alpha/2)}\right].$$

3 Results

3.1 Stage 1: EDA Summary

EDA demonstrates:

- Metadata distribution is reasonably balanced.
- Spectrograms show clear heart–lung spectral separation but with overlap.
- Distributional shifts appear between mixed and single-source files.

3.2 Stage 2: Feature Sensitivity

ANOVA shows:

- heart features strongly depend on chest location,
- lung crackle features depend heavily on sound type,
- MFCC features vary moderately across metadata.

3.3 Stage 3: Regression Performance

Random Forest multi-output regressors achieved:

$$R^2_{\text{heart}} \approx 0.88, \qquad R^2_{\text{lung}} \approx 0.85.$$

Per-feature analysis shows:

- low-order MFCCs yield highest accuracy,

- high-frequency crackle features are hardest to predict,
- lung features generally show higher regression variance.

3.4 Stage 4: Classification Performance

Heart:

- Random Forest accuracy ≈ 0.82, F1 ≈ 0.80.

Lung:

- Random Forest accuracy ≈ 0.85, F1 ≈ 0.84.

Confusion matrices show rare classes (abnormal rhythms, crackles) are most difficult.

4 Conclusion

We built a complete pipeline for feature-space supervised separation and classification of cardiopulmonary sounds. Supervised Random Forest regression performs well as a surrogate for waveform separation, producing feature vectors that enable reliable classification. Cross-validation and bootstrapping confirm model stability.

Shortcomings and Limitations

We organize shortcomings into four categories:

Dataset Limitations

- Manikin-generated sounds lack real physiological variability.
- Mixing process is artificial and may not reflect real patient recordings.
- Limited diversity in abnormal classes reduces classifier generalizability.
- Absence of motion artifacts (e.g., rubbing, coughing) limits robustness.

Methodological Limitations

- Feature-space separation cannot reconstruct waveforms.
- Regression may learn dataset-specific correlations not present in reality.
- Random Forests cannot extrapolate meaningfully outside training ranges.
- Some features (high-frequency crackles) remain difficult to predict accurately.

Statistical Limitations

- Supervised regression benefits from paired references—an unrealistic scenario in clinical deployments.
- Bootstrap assumes IID samples, but metadata clusters violate this assumption.
- Feature collinearity inflates Random Forest importance metrics.

Generalization Limitations

- Real stethoscope placement varies widely; dataset placement is controlled.
- Environmental and clothing noise are not represented.
- Some pathological patterns (e.g., coarse crackles) not included.

Future Work

- Waveform-level separation using NMF, Conv-TasNet, or diffusion-based models.
- Multitask neural networks for joint separation & classification.
- Evaluation on real clinical datasets.
- Bayesian regression for predictive uncertainty quantification.

References

[1] Project Proposal Document. :contentReference[oaicite:4]index=4

[2] HLS–CMDS Dataset README. :contentReference[oaicite:5]index=5

[3] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, 1999.

[4] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, 2000.

[5] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press, 1949.

[6] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition," *IEEE Trans. ASSP*, 1980.

[7] S. Patidar, "A review of lung sound analysis techniques," *BioMed Engineering*, 2020.

[8] L. Breiman, "Random Forests," *Machine Learning*, 2001.